

Published in final edited form as:

Mon Not R Astron Soc. 2018 November ; 481(1): 688–702. doi:10.1093/mnras/sty2303.

From the top down and back up again: star cluster structure from hierarchical star formation

Michael Y. Grudi^{1,★}, Dávid Guszejnov¹, Philip F. Hopkins¹, Astrid Lamberts¹, Michael Boylan-Kolchin², Norman Murray^{3,4}, and Denise Schmitz¹

¹TAPIR, Mailcode 350-17, California Institute of Technology, Pasadena, CA 91125, USA

²Department of Astronomy, The University of Texas at Austin, 2515 Speedway, Stop C1400, Austin, TX 78712, USA

³Canadian Institute for Theoretical Astrophysics, 60 St George Street, University of Toronto, ON M5S 3H8, Canada

⁴Canada Research Chair in Astrophysics

Abstract

Young massive star clusters spanning $\sim 10^4$ – $10^8 M_\odot$ in mass have been observed to have similar surface brightness profiles. We show that recent hydrodynamical simulations of star cluster formation have also produced star clusters with this structure. We argue analytically that this type of mass distribution arises naturally in the relaxation from a hierarchically clustered distribution of stars into a monolithic star cluster through hierarchical merging. We show that initial profiles of finite maximum density will tend to produce successively shallower power-law profiles under hierarchical merging, owing to certain conservation constraints on the phase-space distribution. We perform N -body simulations of a pairwise merger of model star clusters and find that mergers readily produce the shallow surface brightness profiles observed in young massive clusters. Finally, we simulate the relaxation of a hierarchically clustered mass distribution constructed from an idealized fragmentation model. Assuming only power-law spatial and kinematic scaling relations, these numerical experiments are able to reproduce the surface density profiles of observed young massive star clusters. Thus, we bolster the physical motivation for the structure of young massive clusters within the paradigm of hierarchical star formation. This could have important implications for the structure and dynamics of nascent globular clusters.

Keywords

stars; formation–galaxies; star clusters; general–galaxies; star formation

1 INTRODUCTION

Most stars in the Universe are field stars, gravitationally bound only to their host galaxies and not to any discernible smaller element of structure. However, when the locations of

★mgrudich@caltech.edu

initial star formation are considered, there is strong evidence that most stars are born in a statistically clustered, correlated configuration (Lada & Lada 2003; McKee & Ostriker 2007; Bressert et al. 2010; Gouliermis et al. 2015; Grasha et al. 2017; Gouliermis 2018). The star formation efficiency M_{\star}/M_{gas} of typical giant molecular clouds is only of order 1–10 per cent (Myers et al. 1986; Mooney & Solomon 1988; Williams & McKee 1997; Evans et al. 2009; Heiderman et al. 2010; Lada, Lombardi & Alves 2010; Murray 2011; Kennicutt & Evans 2012; Lee, Miville-Deschênes & Murray 2016), possibly due to stellar feedback disrupting the molecular cloud once a certain stellar mass has formed (Murray, Quataert & Thompson 2010; Hopkins, Quataert & Murray 2012; Hopkins et al. 2014; Grudi et al. 2018). The loss of binding energy from the blowout of the remaining gas can unbind the initial stellar distribution (Tutukov 1978; Hills 1980; Mathieu 1983; Lada, Margulis & Dearborn 1984; Elmegreen & Clemens 1985; Baumgardt & Kroupa 2007; Parmentier et al. 2008), allowing most or all stars to disperse into the surrounding galaxy. However, the existence of young, apparently well-relaxed star clusters within the Milky Way (Portegies Zwart, McMillan & Gieles 2010) suggests that a certain fraction of star formation does lead to bound cluster formation, even in Milky Way-like conditions. In many cases, young star clusters have not had time to evolve under the effects of evaporation, dynamical relaxation, and stellar evolution, so their structures should contain some information about their initial formation. A successful model of star cluster formation will be able to clarify this relationship.

In this paper, we discuss the formation of young massive star clusters (YMCs): star clusters that are younger than ~ 100 Myr and more massive than $10^4 M_{\odot}$ (Portegies Zwart et al. 2010).¹ Unlike mature globular clusters, which are generally well-fit by tidally truncated models such as the King (1966) profile, YMCs have been found to have extended power-law profiles with no discernible truncation and hence are better fit by the Elson, Fall & Freeman (1987) surface brightness model (hereafter EFF). This model consists of a core of finite surface brightness μ_0 with an outer surface brightness profile that decays as $\mu \propto R^{-\gamma}$, where γ is the parameter determining the logarithmic slope of the surface brightness profile, hereafter referred to as the ‘profile slope’. If $\gamma < 2$, the integrated stellar mass is divergent, so EFF profiles with $\gamma \sim 2$ are referred to as ‘shallow’, and have a greater proportion of their light in the power-law portion of the surface brightness profile compared to steeper profiles.

YMCs quite often do have shallow profile slopes with γ typically ranging from 2.2 to 3.2 (Elson et al. 1987; Mackey & Gilmore 2003a,b; Portegies Zwart et al. 2010; Ryon et al. 2015), which correspond to 3D density profiles $\rho \propto r^{-3.2}$ – $r^{-4.2}$ in the outer regions. The super star clusters of NGC 7252, despite being three to four orders of magnitude more massive than YMCs of the Local Group, also have profile slopes in this range (Bastian et al. 2013). This agreement across mass scales suggests some scale-free physical mechanism of bound star cluster formation, such that a shallow EFF-like surface brightness profile is generally produced.

¹The definition of Portegies Zwart et al. (2010) also implicitly includes gravitational boundedness, however, we emphasize that the observed YMCs we refer to in this text are not necessarily bound.

One might suppose that the shallow power-law profile of young clusters somehow reflects the initial stellar configuration at the time of star formation, and a smooth cloud of gas turns into a structureless star cluster (e.g. Goodwin 1998). However, observations and simulations (Mac Low & Klessen 2004; McKee & Ostriker 2007; Kruijssen 2013; Krumholz et al. 2014) of star-forming clouds agree that the initial distribution of stellar positions in a star cluster is clumpy and hierarchical, not smooth and monolithic. Thus, presently observed smoothly distributed star clusters are likely to have assembled from a hierarchy of sub-clusters that fragmented out of the parent molecular cloud. If so, the present-day structure of young star clusters is the direct result of top-down fragmentation into stars followed by bottom-up assembly into a single star cluster (see Fig. 1). In this work, we investigate this physical process, arriving at an explanation for the observed structure of YMCs.

This paper is structured as follows: In Section 2, we review observations of the structure of YMCs and compare them to the catalogue of star clusters formed in the Grudi et al. (2018; hereafter Paper I) suite of star cluster formation simulations. We argue that the profile slopes of YMCs are established early in a cluster’s lifetime and hence must emerge from their hierarchical formation events. In Section 3, we discuss how this happens, arguing analytically that the hierarchical merging of sub-clusters generally creates clusters with shallower power-law slopes through phase-space mixing. In Section 4, we test our analytic predictions against N -body simulations of collisionless pairwise star cluster mergers and the collisionless relaxation of a hierarchically clustered mass distribution. In Section 5, we discuss various possible implications and generalizations of our results, and in Section 6, we summarize our main results. Appendix A describes our algorithm for identifying bound star cluster membership from N -body particle data in the simulations of Paper I. In Appendix B, we derive, plot, and provide approximations for various functions that are useful in the analysis of an EFF star cluster model in collisionless equilibrium with arbitrary γ .

2 PROFILE SLOPES OF YMC POPULATIONS

The EFF surface brightness model commonly used to fit YMCs has the form

$$\mu(R) = \mu_{\max} \left(1 + \frac{R^2}{a^2} \right)^{-\gamma/2}, \quad (1)$$

where μ_{\max} is the central surface brightness, R is the projected distance from the centre, a is a scale radius, and γ gives power-law index of the outer brightness profile, hereafter referred to as the ‘profile slope’. The corresponding 3D density profile assuming a constant mass-to-light ratio is

$$\rho(r) = \rho_0 \left(1 + \frac{r^2}{a^2} \right)^{-\frac{\gamma+1}{2}}, \quad (2)$$

where

$$\rho_0 = \frac{M}{a^3} \frac{\Gamma(\frac{\gamma+1}{2})}{3^{1/2} \Gamma(\frac{\gamma-2}{2})} \quad (3)$$

is the central density, M the total mass, a the scale radius, and γ the profile slope. This density profile can be recognized as a generalization of the Plummer (1911) model (corresponding to $\gamma = 4$) to arbitrary profile slope.

Several observed YMC populations are rich enough to be able to discern an underlying distribution of profile slopes. In Fig. 2, we plot the distribution of γ as measured by Ryon et al. (2015) for YMCs in M83, Ryon et al. (2017) for NGC 628 and NGC 131, and Mackey & Gilmore (2003a,b) for the Magellanic Clouds. These clusters range from $\sim 10^6$ – 10^8 yr in age and $\sim 10^4$ – $10^6 M_\odot$ in mass. In all three populations, the median γ is around 2.5. In general, agreement between the observed distributions is quite good, suggesting that a population of EFF-like clusters with this γ distribution arises from some common underlying physical process.

Power-law density profiles have been proposed to emerge in star clusters in various ways. A power-law density profile is the hallmark of gravothermal core collapse, but an inner density profile of $\rho \propto r^{-2.2}$ should generally result (Cohn 1980; Lynden-Bell & Eggleton 1980), which is unlike the outer power-law profile $\rho \propto r^{-3.5}$ typically observed in YMCs. von Hoerner (1957) and Hénon (1964) found that a $\rho(r) \propto r^{-4}$ (hence $\gamma = 3$) density profile results when a uniform collisionless sphere with a Maxwellian velocity distribution undergoes violent relaxation towards collisionless equilibrium. More generally, it results from a discontinuity in the distribution of stellar mass in energy space across the boundary between bound and free orbits, as is caused by the escape of stars with positive energy after a violent relaxation event (Aguilar & White 1986; Jaffe 1987; Merritt, Tremaine & Johnstone 1989). As such, this may be a good model of the initial relaxation of the smallest bound sub-structures, or at the resolution limit in star cluster formation simulations that do not resolve individual stars (e.g. Paper I). However, it does not explain the fact that the majority of star clusters have $\gamma < 3$.

Elson et al. (1987) suggested that the typically observed value $\gamma \sim 2.5$ corresponds to the $\rho \propto r^{-3.5}$ profile found in Spitzer & Shapiro (1972) as a steady-state solution for the outer halo of a star cluster with an inner core, but they proceeded to point out that this structure would have to be established on the two-body relaxation time-scale (Spitzer 1987; Portegies Zwart et al. 2010):

$$t_{\text{rh}} = 4 \times 10^7 \text{ yr} \left(\frac{M}{10^4 M_\odot} \right)^{1/2} \left(\frac{R_{\text{eff}}}{1 \text{ pc}} \right)^{3/2}, \quad (4)$$

where R_{eff} is the half-mass radius (we have also assumed here that the mean mass of a star is $0.5 M_{\odot}$). Many YMCs are much younger than their respective two-body relaxation time-scale, so this picture is not satisfactory.

In general, scenarios requiring more than a few Myr can be ruled out, as good EFF fits appear to have been achieved for quite young star clusters. Indeed, Ryon et al. (2015) found no correlation of γ with cluster age in M83, suggesting that any secular evolutionary processes occurring within these YMCs typically takes longer than ~ 100 Myr to have an appreciable systematic effect on the outer structure. Such young cluster have not existed long enough to experience any significant number of dynamical relaxation times or orbits around the host galaxy during which they may be tidally stripped. Thus, we will explore explanations in which γ is established over a relatively short cluster formation time-scale and then evolves only slowly. The most promising of these is the other physical explanation proposed by EFF: dissipation-less relaxation following a rapid star formation event. It was noted that simulations of the collisionless relaxation of galaxies from a clumpy, non-equilibrium state (van Albada 1982; McGlynn 1984) could reproduce the range of profile slopes observed in star clusters. We will revisit this scenario in the context of modern star formation theory.

2.1 Simulated cluster populations

To guide our analytic exploration, we consider simulations of star cluster formation. The multiphysics N-body magnetohydrodynamic (MHD) simulations of Paper I followed the collapse of a parameter survey of unstable gas clouds with a wide range of initial conditions, e.g. $10\text{--}1000$ pc in diameter and $10^2\text{--}10^4 M_{\odot} \text{pc}^{-2}$ in mean surface density. We found that the clouds form stars until a certain critical stellar surface density has been reached, sufficient to disrupt the cloud via stellar feedback, which included the combination of photoionization heating, radiation pressure, shocked stellar winds, and supernova explosions, approximated numerically according to the methods developed for the FIRE project in Hopkins et al. (2014, 2018). In general, we have found that the simulations with greater star formation efficiency place end with a significant fraction of the total stellar mass in gravitationally bound, virialized star clusters.

These star clusters form via hierarchical assembly,² as has been found in previous simulations following the collapse and turbulent fragmentation of molecular clouds (Klessen & Burkert 2000; Bonnell, Bate & Vine 2003). Many small sub-clusters first fragment out of the molecular cloud, which then go on to merge with their neighbours, eventually building up a massive star cluster. Unlike N-body simulations of star cluster assembly that have relied upon certain assumptions about the mass-loss history of the system (e.g. Scally & Clarke 2002; Fellhauer & Kroupa 2005), the process of star cluster assembly in concert with feedback-induced mass-loss is followed self-consistently by including stellar feedback physics.³

²A visualization of the star cluster formation process can be found at <http://www.tapir.caltech.edu/~mgrudich/gmc.mp4>.

³Unlike these works, our simulations do not resolve the motions of individual stars, however.

We identify and catalogue those star clusters that are both well-resolved (greater than 10^3 particles) and gravitationally bound via the algorithm described in Appendix A. We have found that the surface density profiles of star clusters formed in the simulations are generally well-fit by the EFF profile, covering a range of slope parameters. In Fig. 2, we plot the distribution of slopes extracted from the star cluster populations formed in the simulations of Paper I, both with and without stellar feedback. We find that the agreement with the observed populations is within the observational scatter for the simulations that include stellar feedback, suggesting that at least the most important physics necessary for realistic star cluster structure are accounted for in the simulations. We find no strong correlation between γ and cluster mass, age,⁴ or radius, in agreement with Ryon et al. (2015).

The simulations without stellar feedback also have a significant population of shallow clusters, but there is a deficit of very shallow clusters having $\gamma < 2.5$. Without stellar feedback, the population of bound star clusters tends to be richer: more stars form overall due to the absence of a force that moderates star formation. Also, the clusters are generally denser on average due to the lack of energy input from feedback; they do not undergo dynamical expansion due to mass-loss. These dense, compact clusters are much less likely per orbit to merge with their neighbours, whereas mergers are more common in simulations with feedback because the clusters undergo some amount of dynamical expansion, increasing the cross-section for merging. This suggests that the formation of shallow clusters has something to do with the dynamics of the cluster assembly process.

The above simulations and observations lead us to several hypotheses about the origin of YMC mass profiles:

- (i) The distribution of profile slopes does not differ greatly between different observed or simulated cluster-forming environments, if one accounts for stellar feedback in the simulations.
- (ii) Interactions with the galactic environment are not necessary to reproduce the observed γ distribution, as the simulations do not include these physics.
- (iii) Few-body interactions must play a secondary role in determining the bulk structure of the cluster, as even if the simulations were capable of resolving these effects (which they are not) they do not run for any significant fraction of a half-mass relaxation time. Structural details on the scale of individual stars, such as the stellar mass function, can be neglected in favour of a mean-field, IMF-averaged approximation over time-scales much less than the two-body relaxation time-scale.

It is therefore plausible that star clusters generally form with EFF-like surface brightness profiles, directly from their initial relaxation from their hierarchically clustered state.

⁴Note, however, that these simulations follow the isolated formation of star clusters and do not follow a cluster's subsequent evolution in a galactic environment.

3 SHALLOW CLUSTERS THROUGH MERGING SUB-STRUCTURE

We will now develop physical intuition for how hierarchical star formation leads to the formation of star clusters with shallow power-law profiles. Consider first the initial conditions of the problem: a gas cloud collapses and undergoes star formation. Observations of the M83 YMC population suggest that the majority of the YMCs evacuate their natal gas as soon as 2–3 Myr (Hollyhead et al. 2015), at most a few orbital times. This is also the case in the Paper I simulations. This process of rapid star formation still has some finite duration, but we may consider an idealized model wherein the stars are formed in place instantaneously, and the system then relaxes as a dissipationless N -body system.

This initial arrangement of stars resulting from the fragmentation of the cloud will be hierarchically clustered (e.g. Bonnell et al. 2003; Gouliermis et al. 2015; Grasha et al. 2017; Guszejnov, Hopkins & Krumholz 2017). This is because fragmentation will leave behind sub-structures of all scales from the size of the parent cloud to the scale of protostellar discs (Hopkins 2013). The proportion of the original gas cloud that is actually converted into stars will be limited by the dynamical ejection of gas and the eventual blowout due to stellar feedback (e.g. Murray et al. 2010; Grudi et al. 2018), but let us assume that the cloud has high (>50 per cent) star formation efficiency, which generally leads to the formation of a bound star cluster (Hills 1980; Elmegreen & Efremov 1997). sub-clusters that fragmented from the same self-gravitating parent will then be gravitationally bound to each other on average, so once they have turned into stars they will eventually merge together under dynamical friction. The result will be a sequence of hierarchical merging: sub-clusters will merge with their immediate neighbours that fragmented from the same parent, then the more massive cluster will merge with its neighbour, etc. (see Fig. 1). The smallest and densest structures will merge first because their respective dynamical times are the shortest, as their orbital time will be essentially the free-fall time at the mean density of their parent structure, $t_{\text{ff}} \propto \rho^{-1/2}$.

This process is certainly complex, but the success of the Paper I simulations in producing star clusters with the correct structure out of softened, equal-mass star particles encourages us to consider a collisionless kinetic treatment of the problem. We approximate the dynamics as those of an ensemble of stars with phase-space distribution function $f(\mathbf{x}, \mathbf{v}, t)$, which evolves according to the collisionless Boltzmann equation:

$$\frac{Df}{Dt} = 0, \quad (5)$$

where $\frac{D}{Dt}$ denotes the Lagrangian time derivative along the flow of the system determined by the Hamiltonian with the usual kinetic and gravitational terms. In other words, the phase-space density f is conserved along trajectories of the system. Formally, this does link the initial state of a hierarchical stellar distribution to the final state of a monolithic star cluster. However, it cannot be applied directly: while the fine-grained distribution function f is indeed conserved in a dissipationless relaxation process, the measurable quantity in any observation or N -body simulation is the coarse-grained distribution

$$\bar{f}: \quad \bar{f}(\mathbf{x}, \mathbf{v}, t) = f(\mathbf{x}, \mathbf{v}, t) * K\left(\frac{\mathbf{x}}{\sigma_x}, \frac{\mathbf{v}}{\sigma_v}\right), \quad (6)$$

where K is some 6D smoothing kernel, σ_x and σ_v are the practical resolution limits of position and velocity measurements, and $*$ represents phase-space convolution. In observations and N -body simulations, the finite masses of the bodies impose a mass scale that ultimately determines the practical limit of phase-space resolution: the support of the smoothing kernel must contain a certain number of bodies to be able to convert between the full discrete description and the continuum approximation in any meaningful way.

The collisionless Boltzmann equation does not require that \bar{f} be conserved along phase-space trajectories. To the contrary, in a system relaxing violently towards equilibrium, phase-space elements of varying f tend to be stretched out and tangled together until eventually it is impossible to recover the original value of f at any resolution at which the continuum limit actually applies (Lynden-Bell 1967; Dehnen 2005). The result is a ‘dilution’ of mass in phase-space, wherein \bar{f} will generally decrease. This process is clearly essential in the relaxation of a hierarchically clustered mass distribution into a monolithic cluster, as the initial clumpy state contains more information than the smooth final state, so this information must be effectively lost as mixing entropy. We expect that in collisionless hierarchical cluster assembly dominated by typically equal-mass mergers, violent relaxation should be efficient at driving this phase-space dilution.

The phase-space mixing theorem derived in Dehnen (2005) makes it possible to constrain the evolution of the phase-space distribution in hierarchical merging. Dehnen found that when two collisionless self-gravitating systems merge, the following function of the coarse-grained phase-space density must strictly decrease for all f :

$$D(f) = \int_{\bar{f}(\mathbf{x}, \mathbf{v}) > f} (\bar{f}(\mathbf{x}, \mathbf{v}) - f) d^3x d^3v, \quad (7)$$

which is known as the *excess mass function*. This mixing theorem was used to explain why the inner density profile of a collisionless merger product must have the same slope as the steeper of the progenitors (e.g. Boylan-Kolchin, Ma & Quataert 2005; Kazantzidis, Zentner & Kravtsov 2006). It thus immediately follows that two EFF-like systems must merge into a system with a flat inner density profile.⁵

We can also use the mixing theorem to constrain the outer density profile of the merger. For this purpose, it is more convenient to consider the *reciprocal* excess mass function $M - D(f)$,

⁵In fact, this follows intuitively from the requirement that the maximum phase-space density cannot increase. Systems in virial equilibrium with flat inner profiles have a maximum phase-space density, while systems with power-law inner profiles do not.

where M is the total mass of the system; this quantity must strictly increase during mixing. Dehnen showed that for a system with a 3D outer density profile $\rho \propto r^{-\gamma-1}$,

$$M - D(f) \propto f^{2\gamma-4/2\gamma-1}. \quad (8)$$

For values of γ giving finite mass ($\gamma > 2$), the exponent $f^{2\gamma-4/2\gamma-1}$ increases monotonically from 0 at $\gamma = 2$ to 1 as $\gamma \rightarrow \infty$. Hence, $M - D(f)$ is a steeper function of f for star clusters with steeper outer profiles. Therefore, when two collisionless systems merge, the requirement that the reciprocal mass function for the whole system must increase for all f implies that the function must be at least as shallow as the shallower of the two systems in isolation. Consequently, *the outer density profile of merger product of two collisionless systems can be no steeper than the shallower of the two progenitors*. We are thus able to explain why hierarchical merging does not produce steeper density profiles than existed originally, however, it remains to explain why it might drive the system towards shallower slopes.

3.1 Similarity solution

A shallow outer density profile can be associated with mass being spread over many orders of magnitude in phase-space density. In particular, $dM/d \log \bar{f} \sim \epsilon$, where ϵ is some small fraction of the total mass of the system. More generally, if we consider any parameter describing a ‘scale’ that approaches 0 far away from the system, be it spatial scale, density, phase-space density, or velocity dispersion, it also holds that

$$\frac{dM}{d \log x} \sim \text{const}. \quad (9)$$

for shallow clusters, where x is the chosen scale parameter. In Guszejnov, Hopkins & Grudi (2018), we argue that such a broad distribution of mass across different scales is a general feature of systems formed under the action of gravity and supersonic turbulence, whose equations can be cast in a scale-free form under the physical conditions relevant to star formation. Therefore, $\gamma \sim 2$ is the expected result of hierarchical cluster formation in the limit where the hierarchy of sub-structures covers a large range of scales. In both the fragmentation that produces the hierarchical structure, and the merging that effaces it, the physics can prefer no particular scale, and hence leave a small fraction of the total mass behind at each scale, hence the flat distribution of mass in $\log f$.

This argument predicts $\gamma = 2$ in the limit of cluster formation from a deep hierarchical merger tree; in effect, this is the fixed point for the outer density profile in hierarchical merging. However, clusters with $\gamma > 2$ remain to be explained. Furthermore, we know that some of the simulated star clusters plotted in Fig. 2 do not have particularly extended merger histories; inspection of their merger histories of the least well-resolved clusters considered generally reveals no more than 2–3 major mergers. There is clearly some mechanism that allows clusters to reach shallow slopes with only limited merger histories, which must arise from some change in γ in the pairwise merging of star clusters.

3.2 Shallower density profiles through pairwise merging

Let us idealize hierarchical cluster formation as a sequence of pairwise cluster mergers. By symmetry, such a merger would most typically involve two clusters of similar size, mass, and shape, so we will determine the outcome of a merger of identical star clusters described by EFF profiles with $M = a = 1$ and a particular value of γ . Since the two clusters fragmented out of the same parent under gravitational instability, the two clusters can be expected to be gravitationally bound to each other; for simplicity, we will consider the case in which they collide on a marginally bound parabolic orbit with pericentre smaller enough for the clusters to disrupt each other in one or two passes. In a marginally bound, collisionless merger, mass and energy are approximately conserved (White 1979), so we assume mass and energy are conserved for simplicity. Furthermore, we assume that the merger product is another star cluster with an EFF profile with parameters $M' = 2M$, a' and γ' .

If the merger is homologous ($\gamma' = \gamma$), mass and energy conservation imply that $M' = 2$ and $a' = 2$. Then, the coarse-grained phase-space density $\bar{f} \propto G^{-3/2} M^{-1/2} a^{-3/2}$ in the neighbourhood of an average star is rescaled by $1/4$, which satisfies the constraint that \bar{f} must decrease in the evolution of the system. This ‘uniform mixing’ approximation has proven to be quite predictive in the case of dissipationless elliptical galaxy mergers (Cole et al. 2000; Shen et al. 2003; Boylan-Kolchin et al. 2005; Hopkins et al. 2009). However, the physical nature of phase-space mixing and violent relaxation in elliptical galaxy mergers may well be qualitatively different from star cluster mergers: the cusps of elliptical galaxies are scale-free, so the phase-space dilution factor tends to be roughly constant throughout the system, leading to uniform mixing. Meanwhile star clusters with flat inner profiles do have a characteristic scale imprinted by the maximum density or maximum phase-space density; some memory of the maximum density should persist in the merger.

We make the ansatz, to be justified in Section 4.1, that the maximum phase-space density persists throughout the merger, as phase mixing becomes less efficient as $f \rightarrow f_{\max}$, where f_{\max} is the maximum phase-space density found in either cluster. If so, then γ cannot remain the same while preserving mass and energy, as if it did then f_{\max} would take $1/4$ its original value. Assuming that the merger product is an EFF cluster and conservation of mass, energy, and f_{\max} , we arrive at the following equation for the final cluster’s slope γ' :

$$\mathcal{F}(\gamma') = 2^{5/2} \frac{\mathcal{F}(\gamma)}{\mathcal{W}(\gamma)}, \quad (10)$$

where $\mathcal{W}(\gamma)$ and $\mathcal{F}(\gamma)$ are the dimensionless functions that contain the γ dependence of a cluster’s energy and maximum phase-space density (see equations B9 and B12 for approximate forms and Figs B2 and B4 for plots of these functions). This equation can be solved for γ' numerically. In the case of merging equal mass and size Plummer (1911) models ($\gamma = 4$), the solution is $\gamma' = 2.83$: the final cluster is shallower than its progenitors.

We also consider the ansatz that the central density ρ_0 is conserved. In practice, the predictions of the two ansätze are similar (see Fig. 3). In general, the models predict that

$2 < \gamma' < \gamma$, so a sequence of mergers will drive γ towards a fixed point of 2. Intuitively, mass and energy conservation require the final mass and effective radius to roughly double. This must be achieved without changing the central (phase-space) density significantly, so a shallower slope is required because a shallower cluster has greater central (phase-space) density for a given half-mass radius.

By the arguments above, even very steep ($\gamma \sim 10$) clusters of similar size and mass will merge into a cluster with $\gamma \sim 4$, so only 1–2 major mergers are needed to get a cluster into the interval between 2 and 3 in which most YMCs lie (Fig. 2). As we have established that γ must be established quite early in a cluster's lifetime, this merger history comes from the star cluster's hierarchical assembly process.

4 N-BODY EXPERIMENTS

In the previous section, two claims were made that require verification: that the maximum phase-space density is conserved in a collisionless star cluster merger, and that the sequence of mergers necessary to produce an EFF-like cluster with $\gamma \sim 2$ –3 can arise from the relaxation of a hierarchically clustered stellar distribution. Now, we shall verify these claims with N -body numerical experiments, first of a sequence of pairwise mergers and then of a hierarchically clustered configuration. We use the multiphysics code GIZMO (Hopkins 2015) in a pure N -body configuration. Gravity is solved with a hierarchical BH-tree algorithm derivative of GADGET-3 (Springel 2005). We do not simulate the motion of individual star but rather approximate the solution of the collisionless Vlasov–Poisson equation with a Monte Carlo sampling of the distribution function with equal-mass, softened particles. Throughout, we adopt units such that $G = 1$.

4.1 Pairwise cluster mergers

We first simulate the merger of two Plummer model clusters ($\gamma = 4$) to test the ansätze that their maximum phase-space density should be conserved, and that the end product should be well-fit by an EFF profile with γ given by the solution of equation (10). Once these clusters have merged and the cluster has relaxed to a steady state, we extract this cluster, copy it, and set it up to merge with its copy. To avoid building up a spurious anisotropy along the axis of approach, the orientations of the clusters are randomized between mergers. We repeat this for a total of three simulated mergers. The Plummer-equivalent gravitational softening length is fixed at 0.1 in all runs.

4.1.1 Initial conditions—We construct two Plummer cluster models in collisionless equilibrium, randomly sampling the positions of 125 000 particles per cluster according to the 3D EFF distribution (equation 2) with $M = a = 1$ and $\gamma = 4$. The velocity distribution is assumed to be isotropic and is randomly sampled according to the phase-space distribution function of equation (B11), which is exact for the Plummer model. We find that a single such cluster evolved in isolation for 10^4 half-mass dynamical times has no significant evolution from the Plummer model, so we expect that the particle number is sufficient so that collisional effects play no major role in the merger, which happens after ~ 300 dynamical times. We place the cluster centres 100 length units from each other, with the relative velocity adjusted for a parabolic encounter with a pericentric radius of 1.6, which is

just close enough that the clusters merge in a single pass. We set up the two subsequent mergers in the same way, but we scale the pericentric radius to the half-mass radius of the cluster.

4.1.2 Results—In all simulations, the clusters approach and merge in a single pass after $\mathcal{O}(10)^2$ time units, and by the end of the simulation at $t = 1000$ the new cluster has approached a new collisionless equilibrium. A fraction of the particles are ejected from the system, so the assumption that the final cluster will contain all initial mass and energy does not hold exactly, but the fraction is always <10 per cent. Free particles are deleted from subsequent merger simulations.

Data on the formed clusters are presented in Table 1. We perform EFF fits on the final surface density profiles as projected in three orthogonal different planes. The particle positions are binned into annuli around the centre of the cluster, and we fit the masses within each bin to the EFF model via χ^2 minimization. Since we interpret the particle states as a Monte Carlo sampling of the phase-space distribution, the uncertainty of the mass m in each bin is taken to be the Poisson sampling error m / \sqrt{N} , where N is the number of particles in the bin (valid for sufficiently large N). We find that the EFF model always fits the surface density profiles reasonably well (Fig. 4, panel 1), but not exactly; the reduced χ^2 of the fits are on the order of 100. The clusters are only weakly triaxial, with ellipticity 0.25 at most, so the fit results from different projection planes do not vary greatly. Mergers 2 and 3 both reduce the ellipticity initially created by Merger 1.

We find that the successive mergers do shallow the surface density profiles (Fig. 4, clusters with $\gamma = 4$ merge into $\gamma = 2.69$, then 2.69 into 2.48, and then 2.48 into 2.21). This is not in exact agreement with the analytic predictions of Section 3.2 assuming either conservation of density or phase-space density, however, the analytic and numerical predictions of γ agree within 0.1, and agree upon the general trend of a decrease towards $\gamma = 2$. Perfect agreement with the model is not expected because of the many approximations we have invoked. In particular, it is likely that the obtained slope of 2.69 is shallower than the predicted 2.8 due to the fact that the merger orbit had non-zero angular momentum, which must be redistributed in the final configuration. This would give a mass distribution that is more extended (i.e. with a shallower slope) than a cluster of equal energy with no net angular momentum.

The last assumption of Section 3.2 to be verified is conservation of the maximum phase-space density. We estimate the coarse-grained phase-space density in the neighbourhood of particle i in the most straightforward way, generally known as the pseudo-phase-space density (Taylor & Navarro 2001):

$$\bar{f}_i \propto \frac{\rho_i}{\sigma_i^3}, \quad (11)$$

where $\rho_i = m_i/V_i$ is the density of the particle estimated from its effective volume (Hopkins 2015), and σ_i is the local velocity dispersion computed from the velocities of the particle's 32 nearest neighbours.⁶ In Fig. 4, panel 2, we plot the distribution $dM/d \log f$ and find that indeed, the maximum phase-space density (corresponding to the upper cut-off of the distribution) is conserved from the initial Plummer model to the final merger. Thus, the deviation of γ from analytic predictions is due to the deviation of the phase-space distribution of the cluster from that of an isotropic EFF model. This is evident in Fig. 4: despite the good apparent fits of the surface density of Merger 3 to the EFF model, its distribution of phase-space densities looks quite different from that of an isotropic EFF model in collisionless equilibrium (shown as the dotted line). Rather than having the predicted asymptotic $\propto f^{2\gamma-4/2\gamma-1}$ dependence for small f , the distribution is flat over a finite interval, then falls off steeply above and below that interval. The phase-space density at the lower cut-off corresponds to the mean phase-space density of particles near 100 distance units from the cluster centre, which is the initial separation between the clusters in the merger setup and hence where we expect any scale-free behaviour to break down.

From these results, we may conclude that the assumptions of Section 3.2 were largely valid: the collisionless merger of two EFF clusters fits reasonably well to another EFF cluster, at least in its surface density profile. The profile slope γ is close to that analytically determined by conservation of mass, energy, and f_{\max} ; conservation of mass and energy hold approximately, while conservation of f_{\max} holds exactly, to the extent that can be tested by our noisy estimate of the phase-space density.

4.2 Relaxation of a hierarchically clustered mass distribution

Now, we wish to examine whether a hierarchically clustered distribution of stars with realistic spatial and kinematic scaling relations can form an EFF-like star cluster as it relaxes towards collisionless equilibrium. We arrange particles in such a configuration and simulate their dynamical evolution from the hierarchically clustered state.

4.2.1 Initial conditions—We initialize 64^3 particles in a hierarchically fragmented configuration by recursively bifurcating a population of sub-clusters, starting with a single cluster of unit mass centred at the origin. In each bifurcation, the mass ratio q of the two child fragments is sampled from the lognormal distribution⁷ with $\langle q \rangle = 1$ and $\sigma_{\log q} = 1$. The masses of the fragments are then

$$\begin{aligned} m_1 &= \frac{q}{1+q} m_{\text{parent}}, \\ m_2 &= \frac{1}{1+q} m_{\text{parent}}. \end{aligned} \quad (12)$$

⁶Much more accurate estimates of \bar{f} from N -body data exist (Arad, Dekel & Klypin 2004; Ascasibar & Binney 2005), but the pseudo-phase-space density is suitable for the purposes of this limited analysis.

⁷The choice of a lognormal mass-ratio distribution was arbitrary; we have also run simulations where q is always 1 and have found no major difference in our results.

The relative separation of the fragments \mathbf{x} is sampled from a 3D normal distribution with variance σ_x^2 . We scale σ_x^2 to achieve the desired two-point spatial correlation function $\xi(r) \propto r^{-2}$, where

$$1 + \xi(r) = \frac{\langle n(r) \rangle}{\langle n \rangle}, \quad (13)$$

is the ratio between the average number density of particles in a spherical shell of radius r around a star to the mean stellar number density of the system. $\xi(r)$ quantifies the tightness of the hierarchical clustering at a given scale r . The form $\xi(r) \propto r^{-2}$ matches observations of young star clusters on scales greater than 0.01 pc and is predicted by numerical simulations and general considerations of the scale-free interplay of gravity and supersonic turbulence (Guszejnov et al. 2017, 2018). This scaling is achieved by the ‘isothermal’ scaling $\sigma_x \propto m_{\text{parent}}$, so σ_x is thus determined down to a constant scale factor.

With the separation \mathbf{x} thus sampled, the child clusters are displaced so as to preserve the centre of mass:

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{x}_{\text{parent}} + \frac{1}{1+q} \Delta \mathbf{x}, \\ \mathbf{x}_2 &= \mathbf{x}_{\text{parent}} - \frac{q}{1+q} \Delta \mathbf{x}. \end{aligned} \quad (14)$$

Lastly, the relative velocity \mathbf{v} of the child clusters is sampled from a 3D normal distribution scaled to emulate the $v^2 \propto R$ kinematic relation that is generally observed in GMCs (Larson 1981; Solomon et al. 1987; Bolatto et al. 2008) and is robustly reproduced in simulations of isothermal, self-gravitating turbulent clouds (Kritsuk, Lee & Norman 2013), the idea being that protostars will inherit the kinematics of the interstellar medium from which they formed. This scaling relation is achieved by setting $\sigma_v^2 \propto M^{4/3}$. Then, to conserve momentum

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{v}_{\text{parent}} + \frac{1}{1+q} \Delta \mathbf{v}, \\ \mathbf{v}_2 &= \mathbf{v}_{\text{parent}} - \frac{q}{1+q} \Delta \mathbf{v}. \end{aligned} \quad (15)$$

The bifurcation iteration described by equations 12–15 is applied recursively until the mass of a single particle is reached, so structures exist on all mass scales down to the mass of individual particles. However, recall that these N -body simulations of equal-mass, softened particles are to be interpreted as a Monte Carlo approximation of the solution of the collisionless Boltzmann equation. For this to be valid, any resolved structures should be sampled by a certain number of particles, as biases in the dynamics due to the discreteness of the particles are not part of the desired solution. For this reason, once the clustered

configuration has been generated, we smooth the initial conditions by displacing each particle by a random normally distributed offset with $\sigma = 10^{-3}$; this ensures that structures in the initial conditions are sampled by at least ~ 100 particles. We also set the Plummer-equivalent gravitational softening length to 10^{-3} for consistency (e.g. Barnes 2012).

This procedure generates a clustered particle distribution with the desired spatial and velocity correlations, as shown in Fig. 5. The gravitational binding energy W for this distribution is computed with $G = 1$, and the system is rescaled by a scale factor $1/W$, so that it has unit binding energy. The velocities are scaled to have a total kinetic energy of 0.5, so that the system as a whole has a virial parameter $\alpha = T/W = 0.5$.

4.2.2 Results—We generate three different sets of initial conditions and evolve each system for 35 time units; the unit of time is on the order of the dynamical time-scale of the system.⁸ Within the first few time units, sub-clusters undergo hierarchical assembly into a population of clusters that fly apart from each other and relax into a steady state. The rate-limiting step for the formation of a given cluster is merging time-scale of its last two remaining sub-clusters, which is on the order of their mutual orbital period, at most on the order of several time units.

We identify bound clusters at the end of the simulation via the algorithm described in Appendix A. In general, roughly 80 per cent of particles are found to be gravitationally bound to a cluster, the rest having been dynamically ejected from their original hosts in the violent merging process. The surface density profiles of the clusters are generally well-fit by the EFF model, and we present the fitted γ values in Table 2. The uncertainties quoted in Table 2 include the variation in the γ obtained when projecting the surface density profile in three different orthogonal planes. This variation is generally small compared to the magnitude of γ , as the clusters are only weakly triaxial: their histories of statistically isotropic mergers tend to average away preferred orientations. This is also reflected in the clusters' modest ellipticities, which we also tabulate in Table 2. The ellipticities lie in a similar range to those observed in the LMC cluster population (Frenk & Fall 1982; Kontizas et al. 1989).

It is readily seen from Table 2 that the most massive clusters tend to have γ closer to 2. The initial conditions were smoothed over an effective fixed mass scale M_0 , so a hierarchically assembled cluster of mass M would have to have experienced an effective number of mergers $N = \log_2 \frac{M}{M_0}$, so in these simulations the more massive clusters have experienced more mergers, each of which creates a shallower profile. This anticorrelation between mass and γ should not be interpreted as a prediction of the statistics of actual YMC populations because observed YMCs are the product of many statistically independent star formation events involving physics with only weak dependence on the mass scale (e.g. Fall, Krumholz & Matzner 2010; Guszejnov et al. 2018). In contrast, we have simulated only three different events, all at a single mass scale.

⁸A visualization of Run 2 can be found at <http://www.tapir.caltech.edu/~mgrudich/hierarchical.mp4>.

In summary, these numerical experiments demonstrate that an EFF profile can emerge from the relaxation of a generic, hierarchically clustered mass distribution with power-law spatial and kinematic scaling relations consistent with observations of GMCs and young star clusters.

5 DISCUSSION

5.1 Smooth versus clumpy initial conditions for globular cluster formation

Goodwin (1998) concluded that the assembly of a YMC from an initially clumpy and asymmetric configuration was unlikely, for two main reasons. First, it was found that if the level of initial clumpiness is too great, some sub-clusters can survive for many orbits around the primary assembled cluster. However, Goodwin (1998) simulated the evolution of a collection of clumps with comparable mass and uncorrelated initial positions, not accounting for correlations between sub-cluster positions imprinted by the structure formation process. This problem is averted by a hierarchical configuration, as neighbouring sub-clusters are all but guaranteed to merge. In the numerical experiments of Section 4.2, no persistent satellite clumps were found; the clusters that form tend to do so within a few dynamical times and disperse from each other and within those clusters sub-structure is erased efficiently.

The other problem with clumpy initial conditions noted by Goodwin (1998) was that the ellipticity of the final cluster is sensitive to the flattening of the initial conditions, and essentially any amount of initial flattening produced clusters with ellipticities much larger than have been observed, in the range $[0, 0.28]$ (Kontizas et al. 1989). This problem is averted by the specific hierarchical picture we have considered in this work, wherein mergers at different levels in the hierarchy are uncorrelated in orientation due to an assumed statistical isotropy. From these experiments, we find no cluster with ellipticity greatly exceeding the maximum observed. However, it should be noted that the assumption of statistical isotropy would not necessarily hold if, for example, the initial sub-clusters consisted of ‘beads’ along a filament or a galactic spur. Indeed, it is quite possible that hierarchical star formation does impose large-scale statistical anisotropies. As such, an interesting direction for future work on this problem would be to investigate the effect of physically or observationally motivated anisotropy on hierarchical star cluster assembly. One avenue would be a straightforward modification to our fragmentation model (Section 4.2.1) wherein the directions of the separations x and relative velocities v from one level to another are given a non-zero correlation.

Overall, we find the structure of YMCs to be largely compatible with the paradigm of hierarchical cluster formation that we have considered here. The constraints of Goodwin (1998) upon clumpy initial sub-structure apply to the specific scenario that they simulated, with initial clumps of comparable masses and uncorrelated positions. The nature of the relaxation process appears to be qualitatively different when the initial stellar density and velocity field are initialized in a hierarchical fashion in the manner we have investigated, which takes into account the underlying spatial and kinematic correlation functions observed in star-forming regions.

5.2 Applicability of the collisionless approximation

Throughout this paper, we have approximated the dynamics of the ensemble of stars by assuming that the evolution is collisionless and that stars of different masses are well mixed. Working in this approximation, our N -body simulations represented the stellar distribution as an ensemble of equal-mass, gravitationally-softened particles. This picture is clearly not entirely realistic for star clusters, which are generally dense enough for stellar close encounters to be common enough to affect their long-term dynamical evolution. Bonnell et al. (2003) found that an order-unity fraction of stars have close encounters during hierarchical star cluster formation, so the granularity of stellar mass should clearly have some effect. We expect the collisionless approximation to break down for clusters in which the two-body relaxation time is less than the orbital time, which equation (4) predicts is the case for clusters less massive than $\sim 250 M_{\odot}$. Therefore, we expect the physics considered in this work to be most applicable to the regime of massive star clusters that assembled from sub-clusters more massive than this.

The success of the collisionless approximation in producing star clusters with realistic coarse-grained structure in both multiphysics star cluster formation simulations (Paper I) and the numerical experiments of this paper suggests that it is sufficient for these purposes. The orbital evolution in the hierarchical merging scenario is dominated by rapid changes in the gravitational potential driving violent relaxation, which affects stellar trajectories independently of their mass (Lynden-Bell 1967).

5.3 Star cluster initial conditions

It has become possible in recent years to simulate the direct N -body evolution, and other processes governing the post-formation dynamical evolution, of a globular cluster consisting of as many as $\sim 10^6$ stars (Wang et al. 2016). Such simulations are important for understanding the rich variety of physical mechanisms that caused young star clusters to evolve into present-day mature globulars, but they must assume some initial cluster properties ad hoc. Typically, either the Plummer (1911) or King (1966) model is used as the initial model (Portegies Zwart et al. 2010).

However, since YMCs are well described by the EFF model, and we have given this observation further physical motivation in this paper, we propose that a shallower EFF model is a more realistic initial condition for globular cluster simulations, rather than something that resembles a mature globular cluster. According to the distribution of profile slopes (Fig. 2), a typical model would have $\gamma \sim 2.5$. Compared to a Plummer model of equal mass and half-mass radius, the central density of a $\gamma = 2.5$ profile is more than ten times greater, so collisional effects such as mass segregation and core collapse would likely have much earlier onset.⁹ This could easily mark the difference between runaway core collapse happening before or after the mass-loss and death of massive stars ~ 3 Myr after star formation. This is a critical factor determining whether it is possible for runaway stellar mergers to form a very massive star or an intermediate-mass black hole (IMBH) in the centre of the cluster (Portegies Zwart & McMillan 2002; Gürkan, Freitag & Rasio 2004;

⁹Although they would still take longer than the initial formation of the cluster.

Freitag, Gürkan & Rasio 2006). It should also influence the pairing and hardening of massive stellar binaries centre of dense clusters, which would alter the rate of massive (e.g. $\sim 60 M_{\odot}$) binary black hole mergers such as GW150914 (Rodriguez et al. 2015; Abbott et al. 2016; Rodriguez, Chatterjee & Rasio 2016). Clearly, the detailed early-dynamical evolution of realistic YMC models warrants further study with more realistic initial conditions.

5.4 The outer NFW profile

We have established that the phase-space dilution caused by violent relaxation and phase mixing in the hierarchical merging of star clusters generally drives clusters towards shallower mass profiles approaching $\rho \propto r^{-3}$. Cold dark matter haloes also merge hierarchically and are generally well described by the Navarro, Frenk & White (1996; NFW) profile in cosmological simulations, which also has an r^{-3} dependence. Indeed, it has long been established that such a profile has some relationship with hierarchical merging (White 1979; Villumsen 1982; Duncan, Farouki & Shapiro 1983; McGlynn ; Pearce, Thomas & Couchman 1993). To explain this, we cannot invoke exactly the same argument as the one we have made for star clusters in Section 3 because the NFW model has no maximum phase-space density to conserve. Nevertheless, the Dehnen (2005) mixing theorem still implies that the hierarchical merging of dark matter haloes cannot create steeper density profiles. Furthermore, the outer density profile should behave in a manner that is insensitive to the details of whether the inner profile is a core or a cusp, so shallower density profiles should generally result in mergers. We therefore argue that the $\propto r^{-3}$ outer NFW profile can be understood as the endpoint of the same process of phase-space dilution that we have argued drives star clusters to shallow density profiles.

6 CONCLUSIONS

We arrive at the following conclusions about the formation of young massive clusters:

- (i) We compile observational data of young massive cluster populations (Mackey & Gilmore 2003a,b; Ryon et al. 2015, 2017) and find that the distribution of surface brightness profile slopes (Fig. 2) is similar between different cluster populations, suggesting that it is universal due to common star formation physics.
- (ii) MHD star cluster formation simulations with resolved cooling, fragmentation, and stellar feedback (Grudi et al. 2018) have produced a population of star clusters with profile slopes that agree with observations (Fig. 2), despite the fact that the simulations do not resolve the formation of individual stars. To capture the essential physics that determine the shapes of nascent massive star clusters, it suffices to resolve some fraction of the dynamic range of fragmentation.
- (iii) Stellar feedback clearly has an important role in shaping star clusters, as simulations without feedback are different from observed YMCs in many ways. The role of stellar feedback in setting star cluster structure should be elucidated in detailed cluster formation simulations.

- (iv) Based on the observational and simulation data mentioned above, evidence is strong that a YMC's profile slope is established when it is dynamically young so must be established in the cluster formation process.
- (v) We develop an analytic model for the evolution of a cluster's profile slope γ in a sequence of collisionless pairwise mergers between star clusters modelled by the EFF model. Phase-space mixing requires that the final slope is no shallower than that of either progenitor. Furthermore, assuming conservation of mass, energy, and maximum phase-space density, we find that mergers must always shallow the slope towards 2 by some amount. Thus, a sufficiently large number of hierarchical mergers will result in $\gamma \sim 2$, as argued in Guszejnov et al. (2018) from more general considerations.
- (vi) We perform collisionless N -body simulations of three iterated star cluster mergers, starting with a pair of identical Plummer (1911) models and then merging the result with a copy of itself twice. The results of these simulations are in good agreement with our analytic model: at most ~ 10 per cent of mass and energy are ejected in each merger, the maximum phase-space density is conserved, and the mergers drive γ from 4 initially to a value close to 2 (Table 1). The collisionless merger of two EFF clusters produces another cluster whose surface density profile is also well described by the EFF model, however, deviations from the model are more apparent in the phase-space structure (Fig. 4).
- (vii) We have performed N -body experiments following the collisionless relaxation of a hierarchically clustered mass distribution with spatial and kinematic scaling relations corresponding to those observed in GMCs and young star clusters. We find that sub-clusters rapidly merge hierarchically into steady-state star clusters with EFF-like surface density profiles, despite no initial surface density model being assumed. Thus, the EFF model is physically motivated within the paradigm of hierarchical star cluster formation and indeed EFF's explanation in terms of dissipationless relaxation following rapid star formation is venerated.
- (viii) Because clusters resembling YMCs emerge so readily from plausible star formation physics, a shallow EFF profile is a more plausible model of a nascent star cluster than the commonly simulated Plummer (1911) or King (1966) models. This may have interesting implications for the detailed dynamical evolution of dense star clusters.

ACKNOWLEDGEMENTS

We thank Michael S. Fall, Bruce Elmegreen, Scott Tremaine, and the anonymous referee for helpful feedback. Support for MG and TAPIR co-authors was provided by an Alfred P. Sloan Research Fellowship, NASA ATP Grant NNX14AH35G, and NSF Collaborative Research Grant #1411920 and CAREER grant #1455342. MB-K acknowledges support from NSF grant AST-1517226 and from NASA grants NNX17AG29G and *HST*-AR-13888, *HST*-AR-13896, and *HST*-AR-14282 from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS5-26555. Numerical calculations were run on the Caltech compute clusters 'Zwicky' (NSF MRI award #PHY-0960291) and 'Wheeler' and allocation TG-AST130039 granted by the Extreme Science and Engineering Discovery Environment (XSEDE) supported by the NSF.

APPENDIX A: CLUSTER FINDING ALGORITHM

To identify bound star clusters from the star particle mass, velocity, and position data of the Paper I simulations, we use an algorithm based on identifying potential wells. This is generally more robust than methods based on identifying density maxima because the gravitational potential contains all necessary information for cluster finding, while being inherently smoother and hence less susceptible to noise. The algorithm is as follows:

- (i) Determine some fixed number N_{ngb} of each star particle's nearest neighbours of each star particle's nearest neighbors neighboursin
- (ii) From each particle, move to the neighbour particle with the lowest gravitational potential. Repeat until a local minimum in the potential is found. This is the bottom of the potential well to which the initial particle is now 'associated'.
- (iii) Compute the gravitational potential as sourced only by the particles associated with this potential well in isolation.
- (iv) Associated particles that are bound to the potential well are considered bound members of the cluster.

In practice, we take $N_{\text{ngb}} = 32$, which is the number of neighbour elements used for constructing the hydrodynamic mesh and force softening in the simulations, so it is on the order the size of the least massive self-gravitating structure that can exist in the simulation. A larger value could potentially lump together distinct bound star clusters, while smaller values generally increase the population of spurious clusters. We find this algorithm to have satisfactory accuracy for this problem; it has been tested on control data sets for which the cluster associations are known a priori and stably identifies the same cluster between different simulation snapshots.

APPENDIX B: (SEMI-) ANALYTIC PROPERTIES OF THE EFF MODEL

Here, we derive useful quantities for calculations involving star clusters modelled by the EFF density profile (equation 2) with arbitrary profile slope γ :

$$\rho(r) = \rho_0 \left(1 + \frac{r^2}{a^2} \right)^{-\frac{\gamma+1}{2}}. \quad (\text{B1})$$

The quantities needed to construct a dynamical model with this density profile are only generally expressible in closed form in the special case $\gamma = 4$, which is the Plummer (1911) model. This has ensured its popularity as an initial condition for N -body simulations that is easy to construct. However, as discussed in Section 2, a much more typical initial condition for a star cluster would be $\gamma \sim 2-3$. For those quantities that lack closed-form expressions, we provide approximate expressions or upper and lower bounds for use with numerical root solvers. The reader is also directed to Lupton et al. (1989) for the derivation of the collisionless Jeans model.

B1 Cumulative mass distribution

The cumulative mass distribution for arbitrary γ is

$$\begin{aligned} M(<r) &= \int_0^r 4\pi r'^2 \rho(r') dr' \\ &= \frac{4\pi\rho_0}{3} r^3 {}_2F_1\left(\frac{3}{2}, \frac{\gamma+1}{2}; \frac{5}{2}; -\frac{r^2}{a^2}\right), \end{aligned} \quad (\text{B2})$$

where ${}_2F_1(a, b; c; z)$ is the Gauss hypergeometric function (chapter 15, Abramowitz & Stegun 1965).

B2 Half-mass radius

The 3D half-mass radius R_{eff} may be obtained by solving $M(<r)/M = 1/2$. For the Plummer model ($\gamma = 4$), the solution is $\frac{1+\sqrt{3}}{\sqrt{3}}a \approx 1.3$. For general γ , there is no closed-form solution.

We may derive upper and lower bounds from the constant and power-law parts of the density profile, respectively, from the expansions of $M(r)$ about 0 and ∞ :

$$\left(\frac{3M}{4\pi\rho_0}\right)^{1/3} \leq R_{\text{eff}} \leq \left(\frac{4\Gamma\left(\frac{\gamma+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\gamma}{2}\right)}\right)^{1/\gamma-2} a. \quad (\text{B3})$$

Equipped with these bounds, R_{eff} can be computed efficiently with a bounded root-finding algorithm such as Brent's method. In the limit $\gamma \rightarrow 2$, the solution will approach the upper bound, as most of the mass will be in the power-law portion. Similarly $R_{\text{eff}} \rightarrow (3M/4\pi\rho_0)^{1/3}$ as $\gamma \rightarrow \infty$ because most of the mass will be in the core. In Fig. B1 we plot these bounds in comparison with the true solution.

B3 Potential

The gravitational potential is given by the integral

$$\begin{aligned} \Phi(r) &= \int_\infty^r \frac{GM(r')}{r'^2} dr' \\ &= -\frac{4\pi G a^2 \rho_0 {}_2F_1\left(\frac{1}{2}, \frac{\gamma-1}{2}; \frac{3}{2}; -\frac{r^2}{a^2}\right)}{\gamma-1}. \end{aligned} \quad (\text{B4})$$

The expansion of $\Phi(r)$ about the centre is

$$\Phi(r) = 4\pi G\rho_0 \left(\frac{r^2}{6} - \frac{a^2}{\gamma-1} \right) + \mathcal{O}(r^4). \quad (\text{B5})$$

The shortest possible orbital frequency in the cluster is that associated with simple harmonic motion in the central potential well, which depends only on the central density:

$$\Omega_{\max} = \sqrt{\frac{4\pi G\rho_0}{3}}. \quad (\text{B6})$$

Expanding about $r = \infty$, we see that the leading-order correction to the monopole term $-\frac{GM}{r}$ is

$$\Phi(r) + \frac{GM}{r} \approx \frac{GM}{a} \frac{\Gamma\left(\frac{\gamma-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\gamma}{2}\right)} \left(\frac{r}{a}\right)^{1-\gamma}. \quad (\text{B7})$$

Thus, for larger values of γ , the leading correction to the point mass potential is $\propto r^{1-\gamma}$, which will be very small, so the potential is well approximated by a Keplerian potential. This approximation will be less valid for $\gamma \rightarrow 2$, as most of the mass will be in the power-law portion of the profile.

B4 Energy

A star cluster in dynamical equilibrium will satisfy the virial theorem: $E = -W/2$, where W is the magnitude of the gravitational potential energy. The potential energy associated with the mass distribution may be computed as the integral:

$$W = \int_0^\infty \frac{GM(r)}{r} 4\pi r^2 \rho(r) dr = \mathcal{W}(\gamma) \frac{GM^2}{a}, \quad (\text{B8})$$

where $\mathcal{W}(\gamma)$ is a dimensionless function of γ , plotted in Fig. B2. For the Plummer model, $\mathcal{W}(\gamma) = 3\pi / 32$. The expression in terms of the hypergeometric function is cumbersome, however, it is asymptotically $\propto (\gamma-2)^2$ as $\gamma \rightarrow 2$ and $\propto (\gamma-2)^{1/2}$ as $\gamma \rightarrow \infty$. It can be very well approximated by the following expression:

$$\mathcal{W}(\gamma) = \left(c_1(\gamma-2)^2 + \left(c_2(\gamma-2)^{\frac{1}{2}} \right)^\alpha \right)^{1/\alpha}, \quad (\text{B9})$$

with $c_1 = 0.780$, $c_2 = 0.284$, and $\alpha = -0.692$. This expression interpolates between the two asymptotic behaviours and is indistinguishable from $\mathcal{W}(\gamma)$ as plotted in Fig. B2.

B5 Phase-space distribution function

With the potential given by equation (B4) and assuming an isotropic velocity distribution, the phase-space density $f(\mathbf{x}, \mathbf{v})$ is a function of specific orbital energy alone. We may determine the phase-space density $f(\epsilon)$ with the usual integral formula (Binney & Tremaine 1987):

$$f(\epsilon) = \frac{1}{\sqrt{8}\pi^2} \frac{d}{d\epsilon} \int_{\psi=0}^{\psi=\epsilon} \frac{d\rho}{\sqrt{\epsilon-\psi}}, \quad (\text{B10})$$

where $\psi = -\Phi$ and $\epsilon = (-\Phi - \frac{1}{2}v^2)$. In the limit $r \gg R_{\text{eff}}$, we may approximate $f(\epsilon)$ by substituting the Keplerian potential and the approximation $\rho \sim \rho_0 r^{-\gamma-1}$. In this limit,

$$f(\epsilon) \approx \frac{\Gamma(\gamma+1)\Gamma(\frac{\gamma+3}{2})}{\sqrt{2}\pi^3\Gamma(\frac{\gamma-2}{2})\Gamma(\gamma+\frac{1}{2})} \epsilon^{\gamma-\frac{1}{2}}. \quad (\text{B11})$$

Remarkably, for the Plummer model $\gamma = 4$), this power-law approximation holds exactly. For all other values this is not so, and the integral and derivative in equation (B10) must be taken numerically. The derivative in equation (B10) may be taken with a high-order finite difference stencil, as the integral is smooth everywhere except at $\epsilon = \psi(0)$. Fig. B3 plots the numerically computed $f(\epsilon)$ for various values of γ . It is clear that for the non-Plummer models, the phase-space distribution for the lowest energy (largest ϵ) orbits deviates significantly from a power law. Fig. B4 shows the dependence of the maximum phase-space density upon γ . In the usual units in terms of G , M , and a , the Plummer model has the lowest maximum phase-space density, and with M and a held constant f_{max} increases without bound as $\gamma \rightarrow 2$ and $\gamma \rightarrow \infty$. We may roll the γ dependence into a dimensionless function $\mathcal{F}(\gamma)$, such that $f_{\text{max}} = \mathcal{F}(\gamma) G^{-3/2} M^{-1/2} a^{-3/2}$. An approximation of \mathcal{F} with maximum error $\sim 10^{-4}$ over $\gamma \in [2.01, 10]$ is

$$\mathcal{F}(\gamma) \approx ((c_1(\gamma-2)^{3/4})^\alpha + (c_2(\gamma-2)^{-1/2})^\alpha)^{1/\alpha}, \quad (\text{B12})$$

where $c_1 = 0.0228$, $c_2 = 0.139$, and $\alpha = 0.816$.

B6 Cumulative phase-space-density distribution $M(< f)$

$M(< f)$, the amount of mass at phase-space density less than f , is a useful diagnostic quantity in N -body simulations because it is robust to noisy estimates of f from Monte Carlo particle

data. It is also useful for placing analytic constraints on merger products because it strictly increases in collisionless evolution as phase mixing occurs.

For a spherically symmetric, isotropic cluster model, f is a monotonic function of ϵ , so it is convenient to compute $M(< f)$ as the integral

$$M(< f) = \int_0^{\epsilon(f)} f(\epsilon) g(\epsilon) d\epsilon, \quad (\text{B13})$$

where $\epsilon(f)$ is the inverse function of $f(\epsilon)$ and $g(\epsilon)d\epsilon$ is the phase-space volume within the interval $[\epsilon, \epsilon + d\epsilon]$, computable as

$$g(\epsilon) = \sqrt{2}(4\pi)^2 \int_0^{r(\epsilon)} r^2 \sqrt{\Psi(r) - \epsilon} dr, \quad (\text{B14})$$

where again $r(\epsilon)$ is the radius at which $\Psi(r) = \epsilon$. In the Keplerian approximation, this gives

$$g(\epsilon) \approx \sqrt{2Ma^5} \pi^3 \epsilon^{-5/2}. \quad (\text{B15})$$

Combining this with B11, the asymptotic form of $M(< f)$ is

$$M(< f) \approx \frac{2^{\frac{\gamma-2}{2\gamma-1}} \pi^{\frac{9}{1-2\gamma} + 3} \hat{f} \left(\frac{\hat{f} \Gamma(\frac{\gamma}{2} - 1) \Gamma(\gamma + \frac{1}{2})}{\Gamma(\gamma + 1) \Gamma(\frac{\gamma+3}{2})} \right)^{\frac{3}{1-2\gamma}}}{\gamma - 2} \propto f^{\frac{2\gamma-4}{2\gamma-1}}, \quad (\text{B16})$$

where $\hat{f} = f / (G^{-3/2} M^{-1/2} a^{-3/2})$. In general, the integral B13 must be performed numerically. In Fig. B5, we plot $M(< f)$ for a sequence of EFF clusters with varying γ but equal mass and energy. Note how smaller values of γ have a flatter distribution, so their mass is effectively spread over more orders of magnitude in f .

REFERENCES

- Abbott BP et al., 2016, Phys. Rev. Lett, 116, 061102 [PubMed: 26918975]
 Abramowitz M, Stegun IA, 1965, in Abramowitz M, Stegun IA, eds, Handbook of mathematical functions with formulas, graphs, and mathematical tables. Dover Press, New York
 Aguilar LA, White SDM, 1986, ApJ, 307, 97
 Arad I, Dekel A, Klypin A, 2004, MNRAS, 353, 15
 Ascasibar Y, Binney J, 2005, MNRAS, 356, 872
 Barnes JE, 2012, MNRAS, 425, 1104
 Bastian N, Schweizer F, Goudfrooij P, Larsen SS, Kissler-Patig M, 2013, MNRAS, 431, 1252
 Baumgardt H, Kroupa P, 2007, MNRAS, 380, 1589

- Binney J, Tremaine S, 1987, Galactic Dynamics. Princeton Univ. Press, Princeton, NJ
- Bolatto AD, Leroy AK, Rosolowsky E, Walter F, Blitz L, 2008, ApJ, 686, 948
- Bonnell IA, Bate MR, Vine SG, 2003, MNRAS, 343, 413
- Boylan-Kolchin M, Ma C-P, Quataert E, 2005, MNRAS, 362, 184
- Bressert E et al., 2010, MNRAS, 409, L54
- Cohn H, 1980, ApJ, 242, 765
- Cole S, Lacey CG, Baugh CM, Frenk CS, 2000, MNRAS, 319, 168
- Dehnen W, 2005, MNRAS, 360, 892
- Duncan MJ, Farouki RT, Shapiro SL, 1983, ApJ, 271, 22
- Elmegreen BG, Clemens C, 1985, ApJ, 294, 523
- Elmegreen BG, Efremov YN, 1997, ApJ, 480, 235
- Elson RAW, Fall SM, Freeman KC, 1987, ApJ, 323, 54
- Evans NJ, II et al., 2009, ApJS, 181, 321
- Fall SM, Krumholz MR, Matzner CD, 2010, ApJ, 710, L142
- Fellhauer M, Kroupa P, 2005, ApJ, 630, 879
- Freitag M, Gürkan MA, Rasio FA, 2006, MNRAS, 368, 141
- Frenk CS, Fall SM, 1982, MNRAS, 199, 565
- Goodwin SP, 1998, MNRAS, 294, 47
- Gouliermis DA, 2018, PASP, 130, 072001
- Gouliermis DA et al., 2015, MNRAS, 452, 3508
- Grasha K et al., 2017, ApJ, 840, 113
- Grudi MY, Hopkins PF, Faucher-Giguère C-A, Quataert E, Murray N, Kereš D, 2018, MNRAS, 475, 3511
- Gürkan MA, Freitag M, Rasio FA, 2004, ApJ, 604, 632
- Guszejnov D, Hopkins PF, Krumholz MR, 2017, MNRAS, 468, 4093
- Guszejnov D, Hopkins PF, Grudi MY, 2018, MNRAS, 477, 5139
- Heiderman A, Evans NJ, II, Allen LE, Huard T, Heyer M, 2010, ApJ, 723, 1019
- Hénon M, 1964, Ann. Astrophys., 27, 83
- Hills JG, 1980, ApJ, 235, 986
- Hollyhead K, Bastian N, Adamo A, Silva-Villa E, Dale J, Ryon JE, Gazak Z, 2015, MNRAS, 449, 1106
- Hopkins PF, 2013, MNRAS, 428, 1950
- Hopkins PF, 2015, MNRAS, 450, 53
- Hopkins PF, Lauer TR, Cox TJ, Hernquist L, Kormendy J, 2009, ApJS, 181, 486
- Hopkins PF, Quataert E, Murray N, 2012, MNRAS, 421, 3488
- Hopkins PF, Kereš D, Oñorbe J, Faucher-Giguère CA, Quataert E, Murray N, Bullock JS, 2014, MNRAS, 445, 581
- Hopkins PF et al., 2018, MNRAS, 480, 800
- Jaffe W, 1987, in de Zeeuw PT, ed., Proc. IAU Symp 127, Structure and Dynamics of Elliptical Galaxies. Reidel, Dordrecht, p. 511
- Kazantzidis S, Zentner AR, Kravtsov AV, 2006, ApJ, 641, 647
- Kennicutt RC, Evans NJ, 2012, ARA&A, 50, 531
- King IR, 1966, AJ, 71, 64
- Klessen RS, Burkert A, 2000, ApJS, 128, 287
- Kontizas E, Kontizas M, Sedmak G, Smareglia R, 1989, AJ, 98, 590
- Kritsuk AG, Lee CT, Norman ML, 2013, MNRAS, 436, 3247
- Kruijssen JMD, 2013, preprint (arXiv:1304.4600)
- Krumholz MR et al., 2014, in Beuther H, Klessen RS, Dullemond CP, Henning T, eds, Protostars and Planets VI. Univ. Arizona Press, Tucson, AZ, p. 243

- Lada CJ, Lada EA, 2003, ARA&A, 41, 57
- Lada CJ, Margulis M, Dearborn D, 1984, ApJ, 285, 141
- Lada CJ, Lombardi M, Alves JF, 2010, ApJ, 724, 687
- Larson RB, 1981, MNRAS, 194, 809
- Lee EJ, Miville-Deschênes M-A, Murray NW, 2016, ApJ, 833, 229
- Lupton RH, Fall SM, Freeman KC, Elson RAW, 1989, ApJ, 347, 201
- Lynden-Bell D, 1967, MNRAS, 136, 101
- Lynden-Bell D, Eggleton PP, 1980, MNRAS, 191, 483
- Mac Low M-M, Klessen RS, 2004, Rev. Mod. Phys, 76, 125
- Mackey AD, Gilmore GF, 2003a, MNRAS, 338, 85
- Mackey AD, Gilmore GF, 2003b, MNRAS, 338, 120
- Mathieu RD, 1983, ApJ, 267, L97
- McGlynn TA, 1984, ApJ, 281, 13
- McKee CF, Ostriker EC, 2007, ARA&A, 45, 565
- Merritt D, Tremaine S, Johnstone D, 1989, MNRAS, 236, 829
- Mooney TJ, Solomon PM, 1988, ApJ, 334, L51
- Murray N, 2011, ApJ, 729, 133
- Murray N, Quataert E, Thompson TA, 2010, ApJ, 709, 191
- Myers PC, Dame TM, Thaddeus P, Cohen RS, Silverberg RF, Dwek E, Hauser MG, 1986, ApJ, 301, 398
- Navarro JF, Frenk CS, White SDM, 1996, ApJ, 462, 563
- Parmentier G, Goodwin SP, Kroupa P, Baumgardt H, 2008, ApJ, 678, 347
- Pearce FR, Thomas PA, Couchman HMP, 1993, MNRAS, 264, 497
- Plummer HC, 1911, MNRAS, 71, 460
- Portegies Zwart SF, McMillan SLW, 2002, ApJ, 576, 899
- Portegies Zwart SF, McMillan SLW, Gieles M, 2010, ARA&A, 48, 431
- Rodriguez CL, Morscher M, Pattabiraman B, Chatterjee S, Haster C-J, Rasio F A, 2015, Phys. Rev. Lett, 115, 051101 [PubMed: 26274407]
- Rodriguez CL, Chatterjee S, Rasio FA, 2016, Phys. Rev. D, 93, 084029
- Ryon JE et al., 2015, MNRAS, 452, 525
- Ryon JE et al., 2017, ApJ, 841, 92
- Scally A, Clarke C, 2002, MNRAS, 334, 156
- Shen S, Mo HJ, White SDM, Blanton MR, Kauffmann G, Voges W, Brinkmann J, Csabai I, 2003, MNRAS, 343, 978
- Solomon PM, Rivolo AR, Barrett J, Yahil A, 1987, ApJ, 319, 730
- Spitzer L, Jr., Shapiro SL, 1972, ApJ, 173, 529
- Spitzer L, 1987, Dynamical Evolution of Globular Clusters. Princeton Univ. Press, Princeton, NJ
- Springel V, 2005, MNRAS, 364, 1105
- Taylor JE, Navarro JF, 2001, ApJ, 563, 483
- Tutukov AV, 1978, A&A, 70, 57
- van Albada TS, 1982, MNRAS, 201, 939
- Villumsen JV, 1982, MNRAS, 199, 493
- von Hoerner S, 1957, ApJ, 125, 451
- Wang L et al., 2016, MNRAS, 458, 1450
- White SDM, 1979, MNRAS, 189, 831
- Williams JP, McKee CF, 1997, ApJ, 476, 166

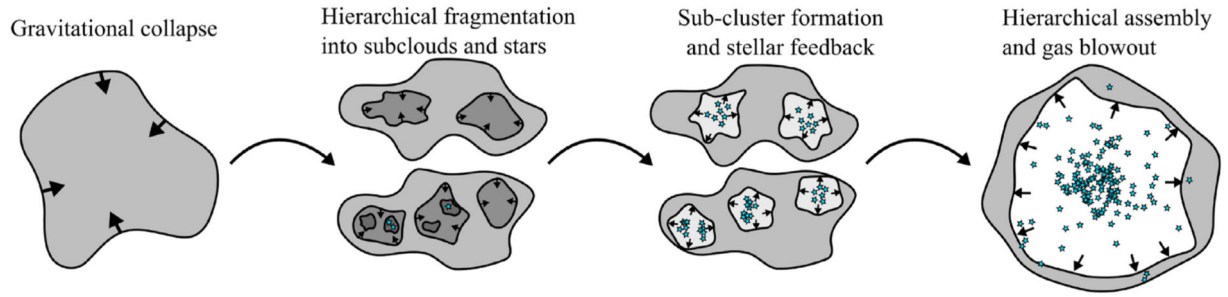


Figure 1.

Proposed model of cluster formation from hierarchical star formation. *Far left:* An unstable molecular cloud undergoes gravitational collapse. *Centre left:* The gravitational instability causes hierarchical fragmentation, producing a hierarchy of sub-clouds that eventually fragment into individual stars. *Centre right:* Stars that fragmented out of the same sub-clouds form in sub-clusters. Feedback from massive stars starts to evacuate gas locally. *Far right:* The sub-clusters merge hierarchically into a single cluster as stellar feedback blows out any remaining gas.

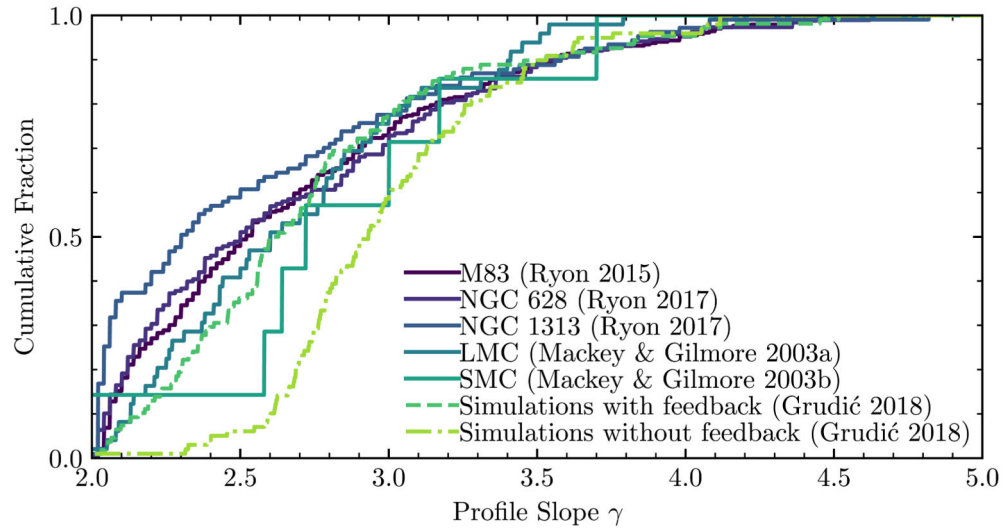


Figure 2.

Solid: Cumulative distribution of star cluster profile slope in the YMC populations of M83 (Ryon et al. 2015), NGC 628, NGC 1313 (Ryon et al. 2017), and the Small and Large Magellanic Clouds (Mackey & Gilmore 2003a,b). *Dashed:* CDF for the star cluster population extracted from the simulations of Paper I, with and without stellar feedback. For both real and simulated cluster populations, we include only those clusters that have $\gamma > 2$, as in Ryon et al. (2015). Agreement between the observed populations is quite good, however, the simulations without feedback appear to have a deficit of shallow clusters. This may be due to the greater compactness of star clusters produced in absence of feedback, which decreases the cross-section for the dynamical interactions that lead to shallower profiles.

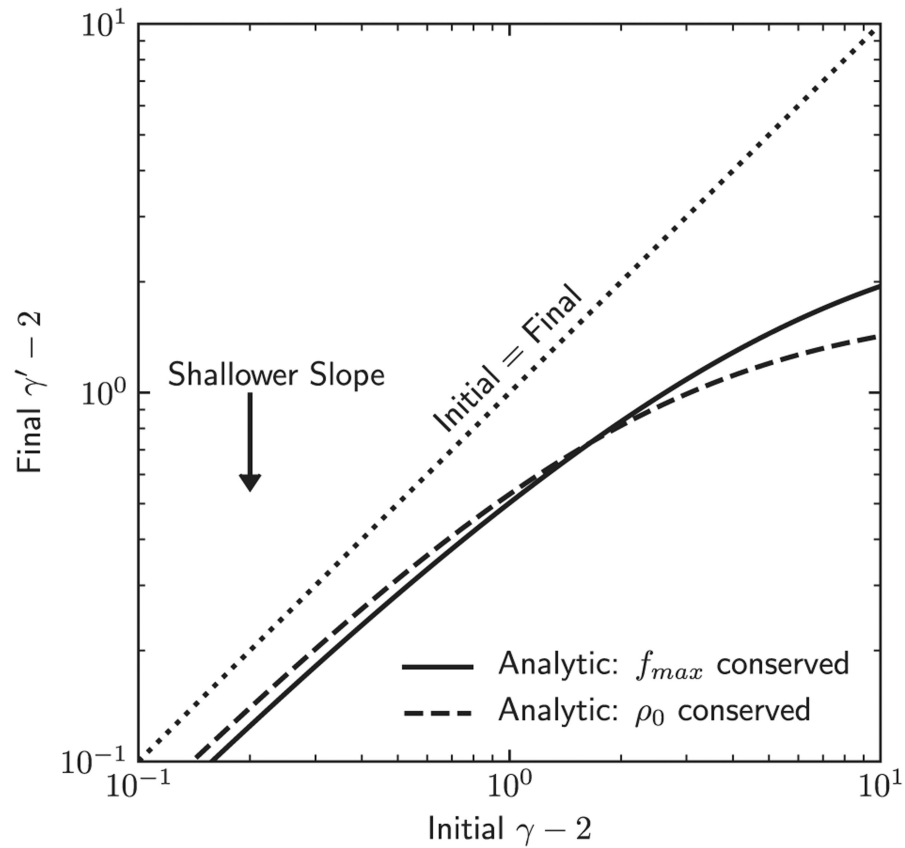


Figure 3.

Final surface brightness slope γ' of the star cluster produced in a merger as a function of the initial γ of two merging clusters with equal γ , mass, and size, assuming that the relaxed merger has an EFF profile. We plot the analytic predictions assuming that the maximum phase-space density f_{\max} (solid) and the maximum density ρ_0 (dashed) are conserved; the two models predict similar results: merging of clusters of equal size and mass always produces a shallower profile than existed before, driving star clusters towards $\gamma = 2$ regardless of their initial structure. We also plot the results of the simulated mergers described in Section 4.1, which do not agree exactly with either model but predict the same overall trend of the formation of shallower profiles.

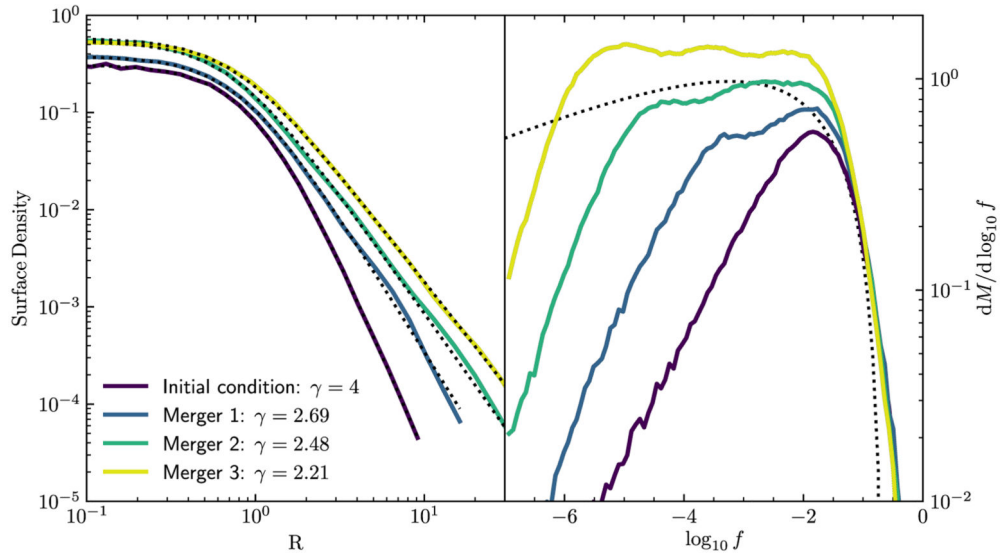


Figure 4.

Results of the successive pairwise merging of star clusters, starting with a pair of identical Plummer models. *Left*: Cluster surface density profiles for the initial Plummer model and the three successive merger products. The mergers generally do shallow the surface density profile towards $\gamma = 2$. Fits to the EFF model are shown as dotted lines. *Right*: Distribution of mass in log phase-space density, $dM/d \log_{10} f$, for the simulated clusters. The mergers generally conserve the maximum phase-space density and distribute the mass across more orders of magnitude in f , gradually building up the flat distribution associated with shallower surface density profiles ($\gamma \sim 2$). The dotted line shows what $dM/d \log_{10} f$ would be for Merger 3 if the phase-space distribution function were that of an isotropic EFF model with fitted parameters fitted from the surface density profile.

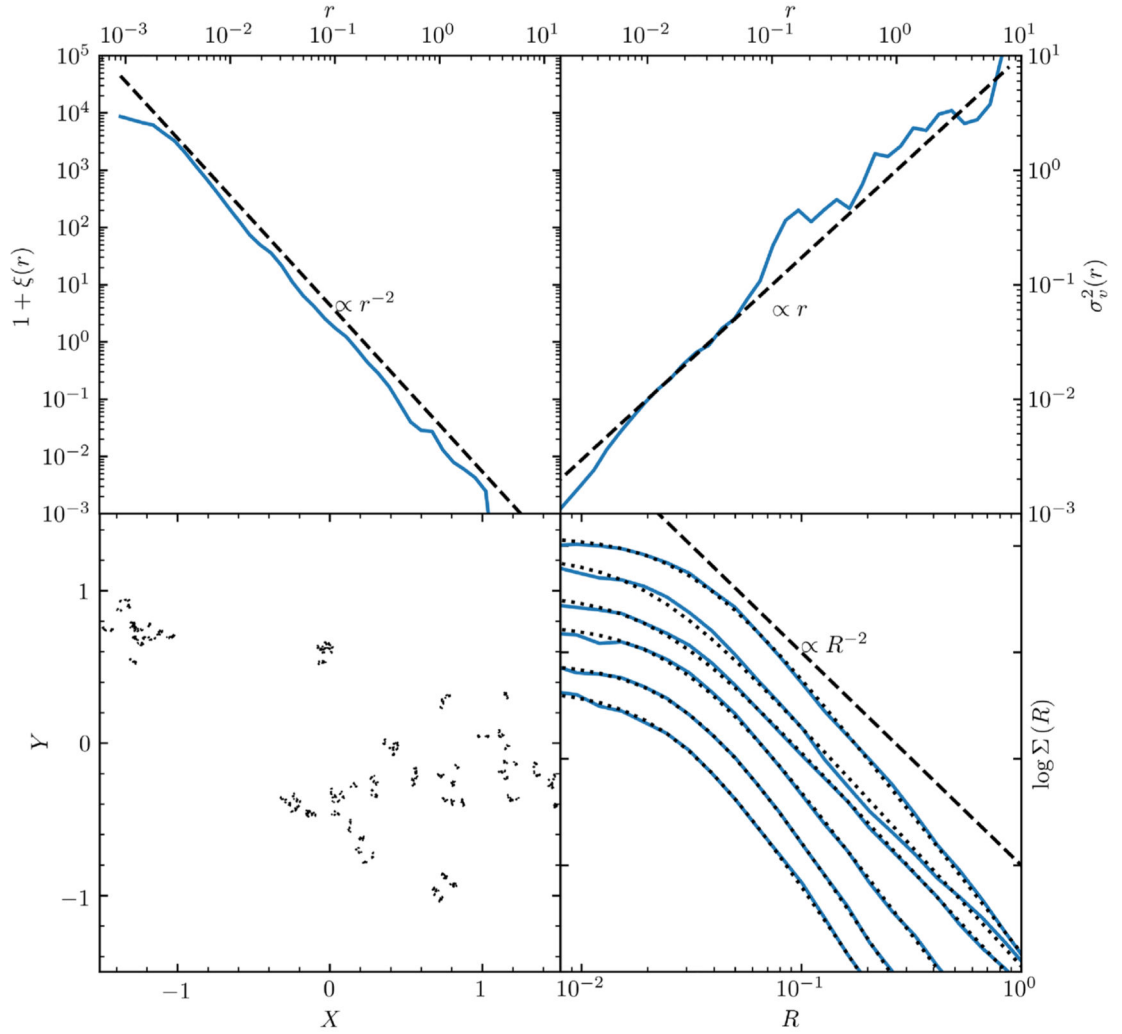
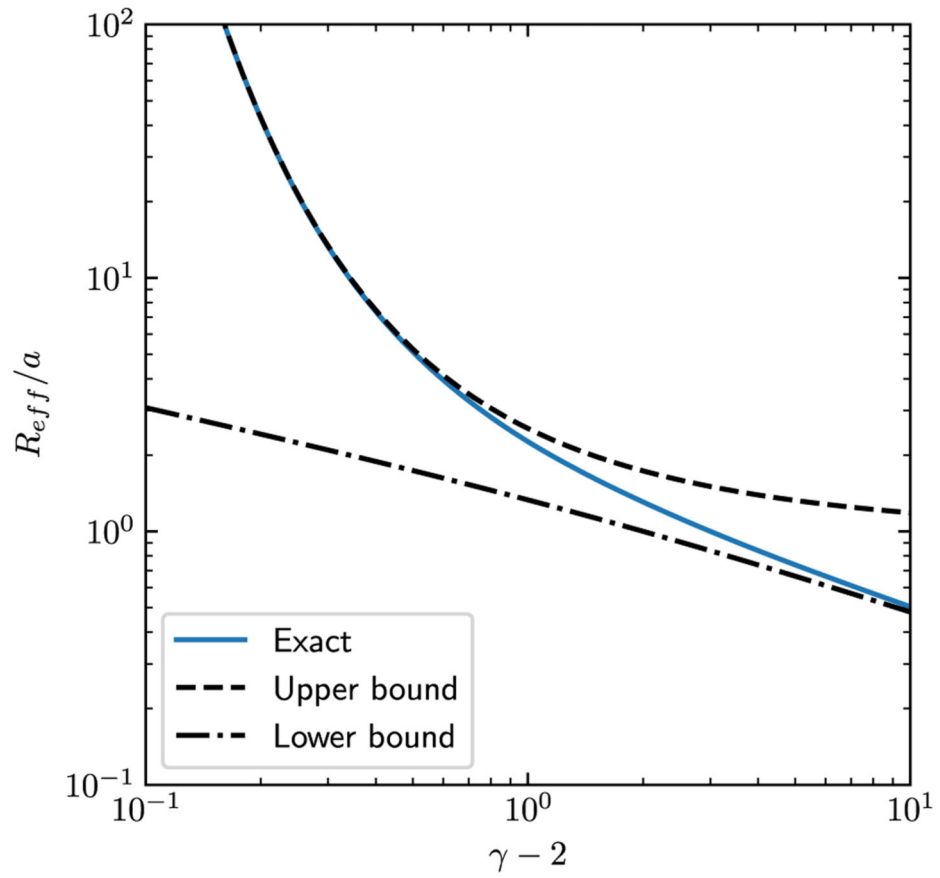
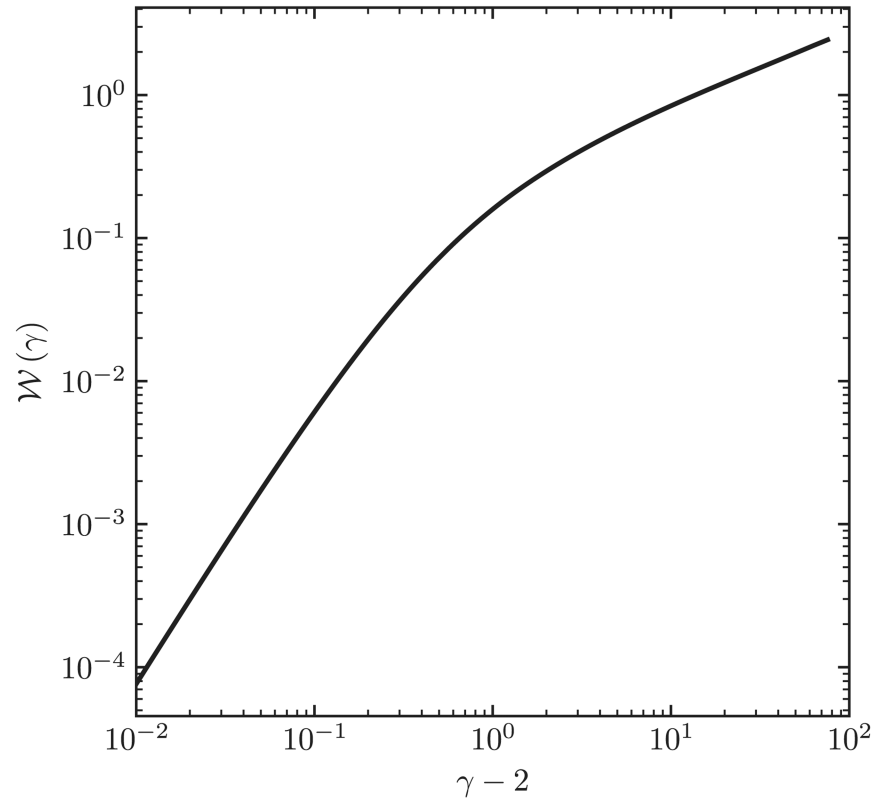


Figure 5.

Initial conditions and final results of a simulation of hierarchical cluster formation, as described in Section 4.2. *Top left:* Initial 3D correlation function of particle positions, which is $\propto r^{-2}$ above the resolution limit. *Top right:* Initial size-velocity dispersion relation. $\sigma_v^2(r)$ is the average velocity dispersion of particles within distance r of any given point and is constructed to be $\propto r$ to agree the observed relation of GMC kinematics (Bolatto et al. 2008). *Lower left:* Initial hierarchically clustered distribution of 64^3 equal-mass particles, constructed by the stochastic fragmentation iteration described in Section 4.2.1. *Lower right:* Surface density profiles of the best-resolved clusters formed by the end of the simulation. The profiles are offset from each other on the plot for visibility. They are well described by the EFF model (equation 1).

**Figure B1.**

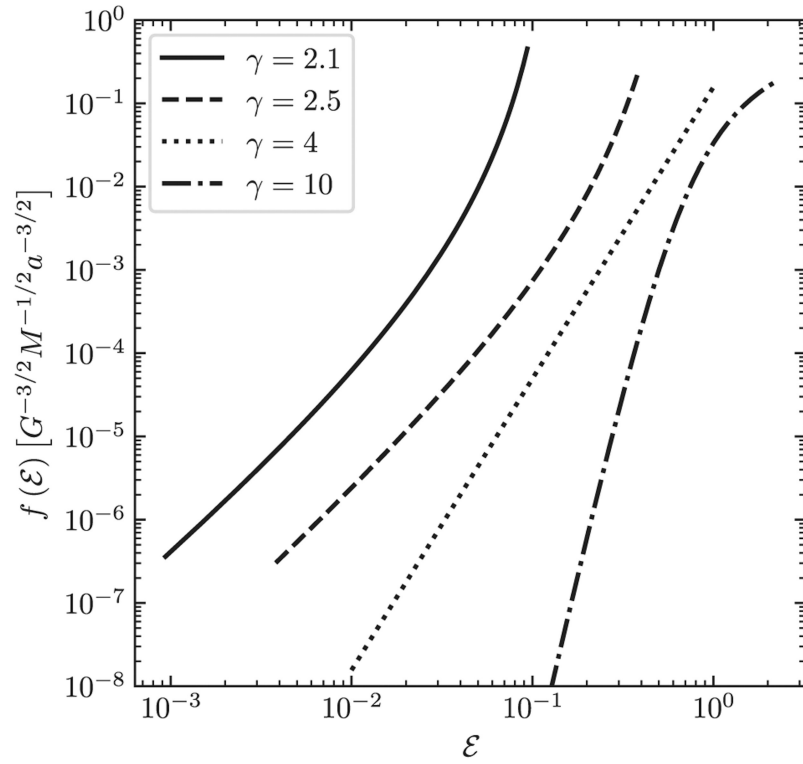
3D half-mass radius R_{eff} as a function of γ in units of the scale radius a . The numerical solution is shown in blue, between the bounds given in equation (B3).

**Figure B2.**

$\mathcal{W}(\gamma)$ as a function of γ , where the gravitational binding energy is given by $W = \mathcal{W}(\gamma) \frac{GM^2}{a}$.

The function is very well approximated by equation (B9). It is $\propto (\gamma - 2)^2$ in the limit $\gamma \rightarrow 2$

and $\propto (\gamma - 2)^{\frac{1}{2}}$ in the limit $\gamma \rightarrow \infty$.

**Figure B3.**

Phase-space density $f(\epsilon)$ in units of $G^{-3/2} M^{-1/2} a^{-3/2}$ for isotropic cluster models with different γ . The Plummer model ($\gamma = 4$) is the only one that is a true power law $\propto \epsilon^{7/2}$, hence its popularity as an analytic model for N -body initial conditions.

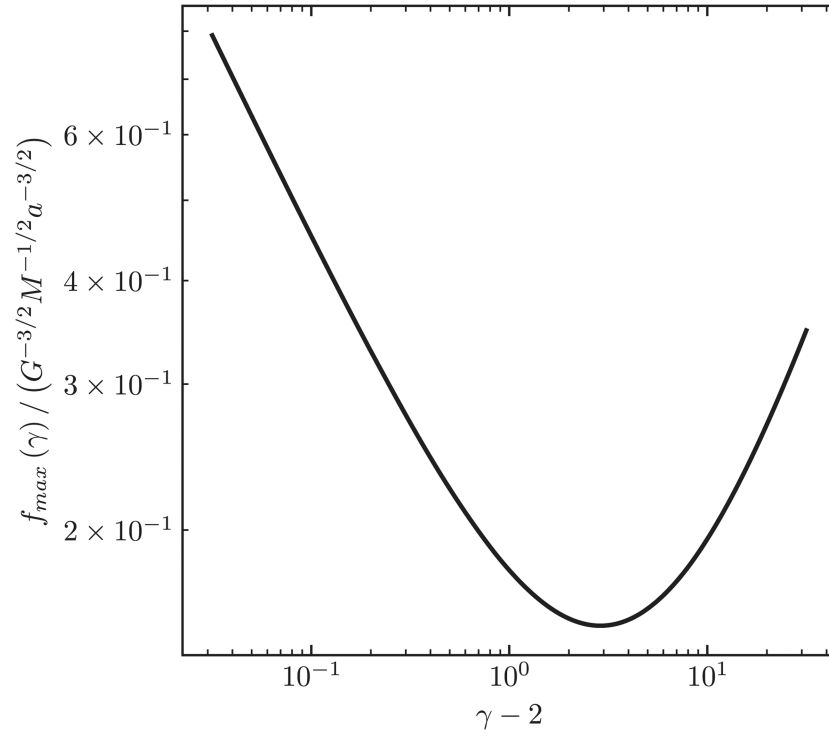


Figure B4.

Maximum phase-space density f_{\max} as a function of γ , in units of $G^{-3/2} M^{-1/2} a^{-3/2}$. The function is $\propto (\gamma - 2)^{-1/2}$ in the limit $\gamma \rightarrow 2$, $\propto (\gamma - 2)^{3/4}$ in the limit $\gamma \rightarrow \infty$ and minimized for the Plummer model ($\gamma = 4$). It is well approximated by equation (B12).

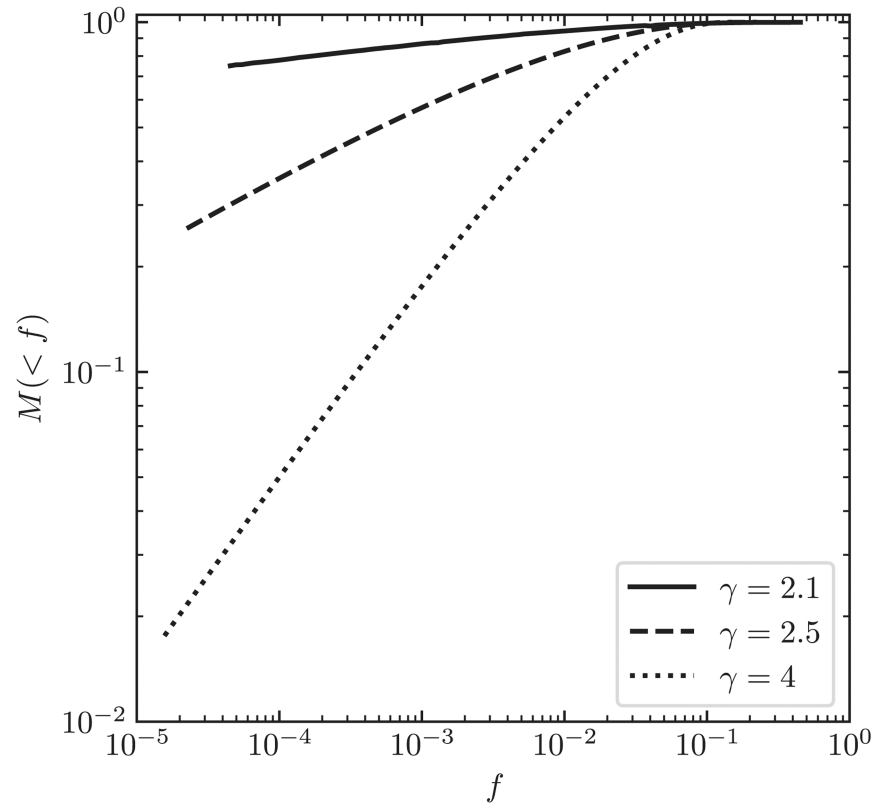


Figure B5.

Cumulative phase-space density distribution $M(< f)$ for a series of clusters varying γ , while keeping mass and energy fixed. At equal-mass and energy, the distribution is more spread-out for γ values closer to 2 and is asymptotically $\propto f^{2\gamma-4/2\gamma-1}$.

Table 1.

Parameters and results of the sequence of simulated mergers of identical EFF-like star clusters: Initial cluster masses M , initial half-mass radii R_{eff} , initial profile slope γ , final cluster mass M' , final half-mass radius R'_{eff} , final fitted profile slope γ' , analytically predicted γ' according to equation (10), cluster ellipticity, and the reduced χ^2 for the fit of the final surface density profile to the EFF model. We give $\hat{\chi}^2_{\text{fit}}$ for the worst of three fits of the final cluster's surface density profile as projected in three orthogonal planes. The quoted uncertainty in γ' includes the variation between the three different fit results.

Run	M	R_{eff}	γ	M'	R'_{eff}	γ'	Predicted γ'	Ellipticity	$\hat{\chi}^2_{\text{fit}}$
Merger 1	1.00	1.30	4.00	1.90	2.24	2.69 ± 0.06	2.83	0.25	78.04
Merger 2	1.90	2.24	2.69	3.57	4.22	2.48 ± 0.03	2.37	0.14	212.24
Merger 3	3.57	4.22	2.48	6.53	7.65	2.21 ± 0.01	2.27	0.13	142.20

Table 2.

Parameters of the clusters produced in the hierarchical relaxation simulations of Section 4.2: Masse, half-mass radius R_{eff} , fitted profile slope γ , ellipticity, and the reduced χ^2 of the surface density fit to obtain γ .

Uncertainties in γ include the variation in the parameters from fitting the surface density profiles as projected in three different orthogonal planes.

Run	Mass	R_{eff}	γ	Ellipticity	$\hat{\chi}_{\text{fit}}^2$
1	0.204	0.108	2.62 ± 0.02	0.13	8.44
1	0.202	0.166	2.26 ± 0.03	0.17	39.87
1	0.195	0.192	2.23 ± 0.02	0.12	13.80
1	0.115	0.074	2.75 ± 0.03	0.15	4.46
1	0.110	0.068	3.16 ± 0.02	0.22	1.12
1	0.054	0.052	3.11 ± 0.04	0.16	1.71
1	0.022	0.038	3.15 ± 0.05	0.12	1.29
1	0.019	0.035	3.19 ± 0.06	0.15	1.31
2	0.382	0.249	2.28 ± 0.03	0.12	104.24
2	0.364	0.171	2.35 ± 0.04	0.17	105.77
2	0.174	0.089	2.89 ± 0.03	0.10	1.78
3	0.147	0.099	2.58 ± 0.02	0.13	4.36
3	0.139	0.083	2.75 ± 0.03	0.21	2.49
3	0.114	0.078	2.62 ± 0.03	0.20	9.19
3	0.106	0.068	2.78 ± 0.03	0.15	6.39
3	0.092	0.067	2.86 ± 0.03	0.26	2.96
3	0.092	0.062	3.22 ± 0.07	0.12	1.54
3	0.053	0.050	3.22 ± 0.04	0.33	1.33
3	0.048	0.051	3.17 ± 0.07	0.25	2.05
3	0.045	0.046	3.48 ± 0.05	0.20	1.28
3	0.043	0.056	2.76 ± 0.03	0.20	2.60
3	0.031	0.040	3.40 ± 0.06	0.13	1.63
3	0.025	0.038	3.32 ± 0.05	0.22	1.10